Title page

Title
Genetic liability to ADHD and substance use disorders in individuals with ADHD

List of authors
Theresa Wimberley, PhD (a,b), Esben Agerbo, DrMedSc (a,b,c), Henriette Thisted Horsdal, PhD (a,b), Cæcilie Ottosen, MD (a,b), Isabell Brikell, PhD (a,b),Thomas Damm Als, PhD (a,d,e), Ditte Demontis, PhD (a,d,e), Anders D. Børglum, PhD (a,d,e), Merete Nordentoft, DrMedSc (a,f), Ole Mors, PhD (a,g), Thomas Werge, DrMedSc,(a,h,i), David Hougaard, DrMedSc (a,j), Jonas Bybjerg-Grauholm (a,j), Marie Bækvad Hansen, PhD (a,j), Preben Bo Mortensen, DrMedSc (a,b,c), Anita Thapar, PhD (k) Lucy Riglin, PhD (k), Kate Langley, PhD (k,l), Søren Dalsgaard, PhD (a,b,c).

Affiliations and adressess

a) iPSYCH - The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Copenhagen and Aarhus, Denmark
b) NCRR - National Centre for Register-based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark
c) CIRRAU - Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
d) Department of Biomedicine and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark
e) Center for Genomics and Personalized Medicine, Central Region Denmark and Aarhus University, Aarhus, Denmark
f) Copenhagen University Hospital, Mental Health Center Copenhagen, Copenhagen, Denmark
g) Psychosis Research Unit, Aarhus University Hospital, Aarhus, Denmark.
h) Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark.
i) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
j) Danish Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
k) Division of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, United Kingdom
l) School of Psychology, Cardiff University, Cardiff, United Kingdom

Corresponding author
Theresa Wimberley, PhD
Aarhus University, National Centre for Register-based Research,
Fuglesangs Allé 26, Building R, 8210 Aarhus V, Denmark.
Email: tw@econ.au.dk Phone: +45 87165976

Running head

Genetic liability and substance abuse in ADHD

Word count

Word count abstract: 303
Word count main text: 3486
Figures: 2 Tables: 3
Supplementary tables: 9
References: 54

Conflicts of interest declaration

None

Abstract

Aims 1) To investigate whether genetic liability to attention-deficit/hyperactivity disorder (ADHD), indexed by polygenic risk scores for ADHD (PRS-ADHD), is associated with substance use disorders (SUD) in individuals with ADHD. 2) To investigate whether other individual- or family-related risk factors for SUD could mediate or confound this association.

Design Population-based cohort study

Setting and participants ADHD cases in the iPSYCH sample, born in Denmark between 1981 and 2003 (N = 13 116). Register-based information on hospital diagnoses of SUD was available until December 31, 2016.

Measurements We estimated odds ratios (ORs) with 95% confidence intervals (CIs) for any SUD as well as for different SUD types (alcohol, cannabis, and other illicit drugs) and severities (use, abuse, and addiction), with effect sizes corresponding to a comparison of the highest PRS-ADHD decile to the lowest.

Findings PRS-ADHD were associated with any SUD (OR = 1.30, 95% CI: 1.11-1.51). Estimates were similar across different types and severity levels of SUD. Other risk factors for SUD (male sex, age at ADHD diagnosis, comorbid conduct problems, and parental factors including SUD, mental disorders, and socio-economic status) were independently associated with increased risk of SUD. PRS-ADHD explained a minor proportion of the variance in SUD (0.2% on the liability scale) compared to the other risk factors. The association between PRS-ADHD and any SUD was slightly attenuated (OR = 1.21, 95% CI: 1.03-1.41) after adjusting for the other risk factors for SUD. Furthermore, associations were nominally higher in females than in males ($OR_{females}$ = 1.59, 95% CI: 1.19-2.12, OR_{males} = 1.18, 95% CI: 0.98-1.42).

Conclusions A higher genetic liability to attention-deficit/hyperactivity disorder was associated with higher risks of substance use disorders in individuals with attention-deficit/hyperactivity

disorder. Results were robust across different types and severity levels of substance use disorders and when adjusting for other risk factors for substance use disorders.

Keywords: attention-deficit/hyperactivity disorder, substance use disorder, polygenic risk, alcohol, cannabis, addiction, predictors, conduct disorder, family history, sex

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable neurodevelopmental disorder with childhood onset, (1) and more frequently diagnosed in males than in females.(2, 3) Among children with ADHD, up to 65% continue to have persistent symptoms and impairments in adulthood, although only about 15% continue to meet full criteria for the disorder.(4)

In addition, children with ADHD are at increased risk of a range of adverse outcomes,(5, 6) including the development of substance use disorders (SUD).(7-9) In general, the risk of SUD increases through adolescence, to reach its maximum during early adulthood,(10, 11) with a faster progression rate in ADHD than in the general population.(12) In individuals with ADHD, known predictors of SUD include persistent ADHD,(13) being diagnosed with ADHD in later adolescence or adulthood,(14) and co-occurrence of oppositional defiant disorder and conduct disorder (ODD/CD).(8, 9, 15) Other factors, including male sex and parental SUD, have been identified as predictors of SUD in the general population in a large US study.(16) A Danish register-based study showed that ADHD increased the risk of SUD less in males than in females.(17)

Recently, the first twelve genome-wide significant common risk alleles for ADHD were identified in the, to date, largest genome-wide association study (GWAS) meta-analysis of ADHD.(18) This meta-analysis included data from the Psychiatric Genomics Consortium (PGC) and the iPSYCH sample – a Danish case-cohort sample of genotyped cases with specific mental disorders, including ADHD, and a 2% genotyped random sample of the Danish population.(19) By including more risk alleles for ADHD, which did not attain genome-wide significance, composite measures of common genetic risk variants can be generated, expressed as polygenic risk scores for ADHD (PRS-ADHD).(18) Using data from the UK Biobank, PRS-ADHD was found to be associated with higher self-reported alcohol intake frequency and dependency in adults without ADHD.(20) A recently published GWAS of cannabis use disorder found PRS-ADHD significantly associated with

cannabis use disorder,(21) and another study recently demonstrated a Single Nucleotide Polymorphism (SNP)-based genetic correlation between ADHD and lifetime cannabis use.(22) Other recently published studies found no strong evidence for an association between PRS-ADHD and SUD.(23, 24) In addition, PRS-ADHD have also been associated with persistence of ADHD(25) and comorbid ODD/CD.(26) Still, there is limited knowledge on whether genetic liability to ADHD is associated with the development of SUD in individuals with ADHD.

To further elucidate the nature of the association between ADHD and SUD and to test whether genetic liability to ADHD is associated with SUD, the present study uses nationwide population-based data available in the iPSYCH sample. The aims were to (1) estimate the association between PRS-ADHD and SUD, across different types and severity levels, in individuals with ADHD; (2) estimate the association between other risk factors and SUD, and (3) estimate the association between PRS-ADHD and SUD while accounting for other risk factors to assess the extent of confounding or mediation.

Methods

Data sources

We linked information from several Danish population-based registers using the unique personal identification number assigned to all individuals living in Denmark and registered in the Danish Civil Registration System since 1968.(27) This register also provides information on sex, date of birth, migration, death, and parents' personal identification numbers.

The Danish Newborn Screening Biobank, contains dried blood spots of practically all infants born in Denmark since May 1, 1981,(28) and for iPSYCH participants provided biological material for genotyping. Genetic information was linked with individual-level register-data. The Danish Psychiatric Central Research Register (DPCRR)(29) and the Danish National Patient Register

(DNPR)(30) providing information on clinical diagnoses according to the International Classification of Diseases, 8th and 10th editions (ICD-8 and ICD-10-Diagnostic Criteria for Research (DCR))(31, 32) and dates of hospital admissions and discharges. Inpatient contacts are registered since 1969 and 1977 in the DPCRR and DNPR, respectively, and outpatient contacts since 1995 in both registers. We obtained information on parental gross income and highest completed education from Statistics Denmark's socioeconomic registers.(33)

Study population

The iPSYCH sample includes a total of 18 726 incident ADHD cases, identified in DPCRR (ICD-10-DCR: F90.0), diagnosed before December 31, 2012.

Biological material was available and passed the genetic quality control for 92% of those with ADHD (N = 17 249).(19) Information on the iPSYCH sample, the genotyping, and quality-control details have been published previously.(18, 19) We restricted our study population to individuals with an incident ADHD diagnosis after age 3, with both parents born in Denmark, born no later than December 31, 2003, alive and living in Denmark at age 13, and with no registered SUD diagnosis before age 13. We further restricted to unrelated individuals, and finally, to reduce population stratification and thereby improve accuracy of PRS-ADHD derived from samples of European ancestry,(34) a genetic homogenous sample was generated by restricting to individuals with European ancestry. This was done by excluding outliers based on a SNP-based principal component analysis (PCA). Methods for restricting to European ancestry and unrelated individuals are described in the supplementary material. Our final study population included 13 116 genotyped individuals with ADHD. Numbers of individuals excluded at each step are provided in Table S1.

Substance use disorders (SUD)

We defined any SUD as at least one ICD-10 diagnosis of SUD (F10-F19) registered after the 13th birthday, in DNPR or DPCRR, regardless of the timing of the first registered ADHD diagnosis. In addition, we categorized SUD into three types: alcohol, cannabis, and other illicit drugs. Finally, we hierarchically ordered severity of SUD into three mutually exclusive levels, based on the first decimal in ICD-10 codes (as shown in Table 2): use (F1x.0), abuse (F1x.1), and addiction (F1x.2), with addiction representing the most severe clinical phenotype. We excluded SUD diagnoses for nicotine use (ICD-10: F17) due to low prevalence (1%), likely due to underreporting of smoking in the registries.

Polygenic risk scores for ADHD (PRS-ADHD)

PRS-ADHD were derived in the iPSYCH sample based on the European ADHD GWAS meta-analysis summary statistics.(18) Scores were calculated as the sum of ADHD risk alleles carried by an individual in the target sample, where each genetic variant was weighted by the corresponding log-odds ratio from the association with ADHD in the discovery sample. Because the Danish iPSYCH ADHD sample constitutes the majority of the entire PGC ADHD sample, scores were calculated using a 'leave-one-out approach'. Cases and controls were split into five groups, and each of these groups were consecutively used as the target sample, using the other four groups plus PGC samples as the discovery sample (details in the supplementary material and elsewhere (18)). Main results are presented for PRS-ADHD at a SNP p-value threshold of 0.2, maximally capturing variance in the original ADHD GWAS.(18)

Finally, we standardized the PRS-ADHD within the five target samples and created deciles.

Risk factors for SUD

We assessed sex (male vs. female) and age at first ADHD diagnosis (≥13 13 years vs. < 13 years) as well as other risk factors for SUD described in the following. Comorbid ODD/CD was defined as a diagnosis of either disorder (ICD-8: 308.03-06, ICD-10: F90.1, F91) in DPCRR or DNPR, from birth and throughout the study period. Parental history of SUD and of any mental disorder was identified if either parent had an SUD diagnosis or a diagnosis of any mental disorder (for ICD codes, see Table S2). Parental socioeconomic status (SES) was defined using previously published definitions.(17) Low paternal income was identified as a gross income in the lowest quintile, based on income for all adult males in the iPSYCH sample in a given calendar year. Low maternal education was defined as having compulsory school education only, usually obtained by age 16. These parental psychiatric and socioeconomic factors were obtained as the last registered information before or during the year the child turned 13 years of age.

Statistical analyses

We estimated odds ratios (ORs) using logistic regression models to examine associations between PRS-ADHD and SUD. The PRS-ADHD were scaled to range from 0 (lowest decile) to 1 (highest decile) and was included as a continuous variable in the primary analyses. One unit increase in PRS-ADHD thus corresponds to a comparison between the highest and the lowest decile. To account for the varying time at risk for SUD, we calculated the observation time as the number of days from age 13 until death, emigration or December 31, 2016, whichever came first. Main analyses were adjusted for observation time, sex, age and calendar year at first ADHD diagnosis (both included as continuous measurements), and the first four principal components (PCs) to adjust for potential remaining population stratification. PCs were obtained from a SNP-based PCA conducted in the final study population (see supplementary material). First, we estimated ORs for the association between PRS-ADHD and SUD. Second, we evaluated other risk factors for SUD by

estimating ORs in independent models. To evaluate the amount of variation in SUD explained by each factor, we additionally estimated Nagelkerke's R^2 (35) and a transformation to the liability scale as proposed by Lee et al. (36) assuming a population prevalence of SUD of 40%.(22, 37) Third, we included each risk factor one by one in the main model, ranked by their estimated amount of variance explained, to investigate whether an association between PRS-ADHD and SUD could be explained by any of these factors. Analyses were repeated for each SUD type and severity level, defined as binary outcome variables.

Moreover, we tested for interactions between PRS-ADHD and the other risk factors for SUD. For each analysis, interactions with other covariates were included in the model.(38) We repeated the main analyses in males and females, separately. Finally, to evaluate each risk factor as a potential confounder of mediator, we assessed their association with PRS-ADHD, by calculating mean PRS-ADHD differences between levels of each risk factor for SUD, and evaluated these by t-tests.

All estimates are accompanied by 95% confidence intervals (CIs), and estimates were declared statistically significant at the 5% level, if not otherwise stated. R^2 and liability estimates were accompanied by non-parametric 95% CIs using bootstrap resampling with 10 000 replications, estimated in SAS version 9.4 (SAS Institute Inc., Cary, NC). The main analyses were conducted using Stata version 15.(39) The PCs used for outlier removal were calculated in EIGENSOFT version 6.1.4(40) and relatedness estimation and final PCA were conducted in PLINK version 1.9.(41-43)

Sensitivity analyses

To check robustness of results, we repeated the main analysis with PRS-ADHD based on different p-value thresholds for SNPs included. In addition, we used different values of assumed population prevalence of SUD in ADHD for the transformation of R^2 into the liability scale. Finally, we

conducted sensitivity analyses to evaluate the impact of applying different cohort restrictions in terms of genetic homogeneity, timing and risk of diagnoses (see supplementary material for further description).

Results

Among the 13 116 individuals with ADHD, 9 680 were males (73.8%) and 3 436 were females (26.2%). Median age at first ADHD diagnosis was 11.5 years (inter-quartile range (IQR) 8.4-17.3), with higher ages in females (median 15.1 years, IQR 9.7-19.8) than in males (median 10.7 years, IQR 8.1-15.9). In total, 2 368 (18.1%) individuals with ADHD developed SUD: 1 731 of these were males (73.1%) and 637 were females (26.9%). Median age at first SUD diagnosis was 19.4 years (IQR 17.2–22.3 years). Characteristics of the study population are presented separately for ADHD individuals with and without SUD (Table 1).

PRS-ADHD and SUD

In the main analyses, adjusted for observation time, sex, age and calendar year at first ADHD diagnosis, and the first four PCs, we found an association between PRS-ADHD and risk of any SUD (OR = 1.30, 95% CI: 1.11-1.51). PRS-ADHD explained approximately 0.2% of the variation in SUD on the liability scale (Table 2). Figure 1 shows the ORs for any SUD by deciles of PRS-ADHD, with the lowest decile as the reference.

Across different types of SUD, associations were observed for alcohol, OR = 1.26 (95% CI: 1.04-1.53) and cannabis, OR = 1.34 (95% CI: 1.10-1.64). Across different levels of severity, associations were observed for substance use, OR = 1.36 (95% CI: 1.02-1.80), and addiction, OR = 1.30 (95% CI: 1.07-1.57). Associations with other illicit drugs, OR = 1.21 (95% CI: 0.99-1.50) and SUD

abuse, OR = 1.21 (95% CI: 0.88-1.65) were less robust with confidence intervals crossing 1 (Figure 2 and Table 3).

Influence of other risk factors for SUD

The ORs with 95% confidence intervals across all SUD outcomes for individual associations with the different risk factors for SUD are illustrated in Figure 2 and presented in Table S3.

The factors most strongly associated with SUD, across types and severity levels, were higher age at first ADHD diagnosis, OR = 3.28 (95 % CI: 2.75-3.92) and comorbid ODD/CD, OR = 2.87 (95% CI: 2.53-3.26). Comorbid ODD/CD was the factor explaining the largest proportion of variation in SUD, as estimated by Nagelkerke's R^2 (3.6%) and on the liability scale (5.6%) (Table 2). All SUD risk factors, except sex and age at ADHD diagnosis, were associated with PRS-ADHD, as estimated by differences in mean PRS-ADHD and p-values evaluated according to a Bonferroni-corrected p-value threshold of 0.007, (Table S4). No evidence of interaction with PRS-ADHD were found for any of the risk factors (Table S5). Associations between PRS-ADHD and any SUD were somewhat attenuated, but remained, after adjusting for the other SUD risk factors (Table 3). Similar results were observed across SUD severity and subtypes, although CI's included 1 in some adjusted models.

Sex-specific analyses

In analyses stratified by sex, we found the strongest association between PRS-ADHD and SUD in females (OR = 1.59, 95% CI: 1.19-2.12), and less robust association in males (OR = 1.18, 95% CI: 0.98-1.42). For specific types and severity levels of SUD, the conclusions were overall the same. However, in males, we found some evidence for an association with cannabis use, OR = 1.26 (95%

CI: 1.01-1.60). In females, we observed no association with alcohol, OR = 1.22 (95% CI: 0.86-1.73), but an association with other illicit drugs, OR = 1.78 (95% CI: 1.18-2.69) (Table S6).

Sensitivity analyses

When we repeated the main analysis with PRS-ADHD based on different p-value thresholds, the effect sizes were largely unchanged across different thresholds (Table S7). Estimated proportion of variance of SUD on the liability scale was largely the same regardless of SUD population prevalence assumed (k=35%, 40%, 45%) (Table S8). The variance explained was slightly increased when including the original continuous and standardised PRS-ADHD instead of the categorized PRS-ADHD in the main analysis (0.26% vs. 0.23%).

Moreover, results were robust to PRS-ADHD based on the iPSYCH sample in the discovery sample only, across different adjustments for genetic population structure, and across different cohort selections (Table S9).

Discussion

This is the first study to examine whether PRS-ADHD is associated with SUD in ADHD, with several novel findings. First, we found that a higher PRS-ADHD increased the risk of any SUD, as well as different SUD types and severities. However, we did not find evidence for association with other illicit drugs, and we found no hierarchy of associations with increasing level of SUD severity. Second, we present novel findings on other risk factors for SUD in ADHD, including late diagnosis of ADHD (after age 13 years), known SUD or mental disorders in parents, and low parental SES, which were all individually associated with increased risks of SUD in individuals with ADHD. In comparison, the PRS explained a much smaller proportion of the variance in SUD. Third, we found that the association of PRS-ADHD with SUD existed over and above what could be explained by

the other risk factors, and was mainly driven by cannabis use. Finally, we observed stronger associations in females. For both males and females, we found an association between PRS-ADHD and cannabis use disorder.

ADHD is a highly heritable disorder and yet only a few studies have examined whether genetic liability to ADHD is associated with an increased risk of SUD, and to our knowledge, no previous study has investigated this association among individuals with ADHD. Our results corroborate the recent finding of a genetic correlation between ADHD and lifetime cannabis use disorder, based on summary statistics from GWAS meta-analyses.(21, 22) However, a case-control study in a Spanish population found no association between PRS-ADHD and SUD as estimated by R^2, probably due to lack of power.(24) Additionally, a study of 1 050 African Americans found no evidence of an association between PRS-ADHD and marijuana use disorders.(44)

Our finding of stronger associations between PRS-ADHD and SUD in females than in males is intriguing and requires further research. Some previous observational studies of ADHD samples suggest higher relative risks of SUD and other adverse outcomes in females than in males,(5, 17) probably explained by females with ADHD often being perceived as less impaired than males,(45) and the higher referral threshold for females.(46) Hence, ADHD severity (and not only the presence of an ADHD diagnosis) may increase the risk of SUD more in females than in males, as indicated by the present study. Despite strong evidence for associations between ADHD and SUD, little is known regarding risk factors for this. We identified, in addition to PRS-ADHD, several other risk factors for SUD in ADHD (e.g. late diagnosis of ADHD and parental history of SUD, mental disorders, and low SES), some of which are known to be associated with SUD in the general population.(47, 48) Comorbid ODD/CD was associated with higher PRS-ADHD and was the

strongest risk factor for SUD, in terms of the amount of variance explained. Still, no sufficient evidence of an interaction with PRS-ADHD was found, and the association we found between PRS-ADHD and SUD, was not fully explained by the presence of comorbid ODD/CD.

This was a nationwide population-based study on the largest genotyped ADHD-sample, based on information on clinical diagnoses of mental disorders made by psychiatrists. However, our study also had several limitations. First, only diagnoses from hospital contacts are included in the Danish registers and SUD is underreported by clinicians.(49) Hence, the prevalence of SUD is likely underestimated and our study may only include a smaller proportion of those with the lowest severities of SUD (use and abuse). In addition, the possibility of misclassification of ADHD and SUD cannot be excluded. Still, the validity of clinical ADHD and SUD diagnoses in the registers has been shown to be high.(50, 51) Furthermore, the prevalence of ADHD diagnoses and treatment in Denmark is low.(52) In order to include as many ADHD cases as possible, and based on the hypothesis of a shared genetic liability increasing the risk of both disorders, we did not take the order of diagnoses into account. This may have influenced our estimation of the effect of age at ADHD diagnosis on risk of SUD. However, in our sensitivity analyses of individuals diagnosed with ADHD before age 13, the association between older age at ADHD diagnosis and SUD remained, as did the association between PRS-ADHD and SUD. Similarly, PRS-ADHD predicted SUD in cases with no SUD before first ADHD diagnosis. Second, collider bias is a fundamental concern in progression studies.(53) More specifically, by conditioning on ADHD, a spurious correlation between the PRS-ADHD and SUD can be introduced through factors such as sex, that both affect ADHD onset as well as SUD. In this particular case, this spurious correlation is removed by adjusting for sex. In the present study, we only assessed genetic liability to ADHD on the risk of SUD. Future research should include measures of other genetic liabilities, such as genetic liability

to SUD and other mental disorders, to evaluate whether these are more or less predictive of SUD. Other limitations were that the present study design did not enable us to account for use of medication and that we were unable to study nicotine use disorder due to very few diagnoses in the registers, as smoking is underreported. In addition, other imprecisions in our register-data may explain part of the association between genetic liability to ADHD and SUD, e.g. depression diagnosed during childhood or imperfect classification of SES by parental education and income levels. Finally, the variance in SUD explained by PRS-ADHD was small, also compared to other individual and family-related risk factors. This is in line with previous research demonstrating that the trait variance in ADHD status explained by the PRS is modest, and the variance explained for secondary outcomes is even lower.(20) PRS prediction may be improved by including larger sample sizes and other genetic factors such as copy number and rare as well as intermediate frequency variants.(18, 54)

In conclusion, this is, to our knowledge, the first study to find evidence that genetic liability to ADHD, indexed by PRS, is associated with higher risk of SUD in individuals with ADHD. Male sex, higher age at ADHD diagnosis, comorbid ODD/CD, parental factors (SUD, mental disorders, low paternal income, and low maternal education) were also identified to be associated with SUD in ADHD. The association between a genetic liability to ADHD with respect to common variants and SUD existed over and above what could be explained by other risk factors for SUD.

Acknowledgements

This study was funded by grants from Novo Nordisk Foundation (NNF16OC0022018), The Lundbeck Foundation (iPSYCH grant no R102-A9118 and R155-2014-1724), and The Stanley Medical Research Institute, and supported by CIRRAU. Dr. Dalsgaard's research is further

supported by grants from Aarhus University Research Foundation (AUFF-E-2015-FLS-8-61), National Institute of Health (R01, grant no ES026993), and the European Commission (Horizon 2020, grant no 667302). The Wellcome Trust provided additional funding for LR and AT (204895/Z/16/Z). Data handling and analysis on the GenomeDK HPC facility was supported by NIMH (1U01MH109514-01 to ADB) and Center for Genomics and Personalized Medicine (grant to ADB). Dr. Børglum's research was further supported by the European Community (EC) Horizon 2020 Programme (grant 667302 (CoCA)).

References

1. Thapar A, Cooper M, Eyre O, Langley K. What have we learnt about the causes of ADHD? J Child Psychol Psychiatry. 2013;54(1):3-16.
2. Martin J, Walters RK, Demontis D, Mattheisen M, Lee SH, Robinson E, et al. A Genetic Investigation of Sex Bias in the Prevalence of Attention-Deficit/Hyperactivity Disorder. Biol Psychiatry. 2017.
3. Novik TS, Hervas A, Ralston SJ, Dalsgaard S, Rodrigues Pereira R, Lorenzo MJ. Influence of gender on attention-deficit/hyperactivity disorder in Europe--ADORE. Eur Child Adolesc Psychiatry. 2006;15 Suppl 1(S1):I15-24.
4. Faraone SV, Biederman J, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. Psychol Med. 2006;36(2):159-65.
5. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. Lancet. 2015;385(9983):2190-6.
6. Dalsgaard S, Mortensen PB, Frydenberg M, Thomsen PH. Long-term criminal outcome of children with attention deficit hyperactivity disorder. Crim Behav Ment Health. 2013;23(2):86-98.
7. Wilens TE. Attention deficit hyperactivity disorder and substance use disorders. A J Psychiatry. 2006;163(12):2059-63.
8. Wilens TE, Morrison NR. The intersection of attention-deficit/hyperactivity disorder and substance abuse. Current opinion in psychiatry. 2011;24(4):280-5.
9. Dalsgaard S, Mortensen PB, Frydenberg M, Thomsen PH. ADHD, stimulant treatment in childhood and subsequent substance abuse in adulthood - a naturalistic long-term follow-up study. Addict Behav. 2014;39(1):325-8.
10. Pedersen CB, Mors O, Bertelsen A, Waltoft BL, Agerbo E, McGrath JJ, et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA Psychiatry. 2014;71(5):573-81.

11. Dalsgaard S, Thorsteinsson E, Trabjerg BB, Schullehner J, Plana-Ripoll O, Brikell I, et al. A comprehensive nationwide study of the incidence of the full spectrum of diagnosed mental disorders in childhood and adolescence. JAMA Psychiatry. 2019;In Press.
12. Molina BSG, Howard AL, Swanson JM, Stehli A, Mitchell JT, Kennedy TM, et al. Substance use through adolescence into early adulthood after childhood-diagnosed ADHD: findings from the MTA longitudinal study. J Child Psychol Psychiatry. 2018.
13. Huntley Z, Young S. Alcohol and substance use history among ADHD adults: the relationship with persistent and remitting symptoms, personality, employment, and history of service use. J Atten Disord. 2014;18(1):82-90.
14. Faraone SV, Wilens TE, Petty C, Antshel K, Spencer T, Biederman J. Substance use among ADHD adults: implications of late onset and subthreshold diagnoses. The American journal on addictions. 2007;16 Suppl 1:24-32; quiz 3-4.
15. Molina BS, Pelham WE, Jr. Childhood predictors of adolescent substance use in a longitudinal study of children with ADHD. J Abnorm Psychol. 2003;112(3):497-507.
16. Blanco C, Florez-Salamanca L, Secades-Villa R, Wang S, Hasin DS. Predictors of initiation of nicotine, alcohol, cannabis, and cocaine use: Results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The American journal on addictions. 2018;27(6):477-84.
17. Ottosen C, Larsen JT, Faraone SV, Chen Q, Hartman C, Larsson H, et al. Sex Differences in Comorbidity Patterns of Attention-Deficit/Hyperactivity Disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 2019;58(4):412-22 e3.
18. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51(1):63-75.
19. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Baekvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2018;23(1):6-14.
20. Du Rietz E, Coleman J, Glanville K, Choi SW, O'Reilly PF, Kuntsi J. Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2017.
21. Demontis D, Rajagopal VM, Als TD, Grove J, Pallesen J, Hjorthøj C, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. Nature Neuroscience (in Press) bioRxiv 2018. 2019.
22. Soler Artigas M, Sanchez-Mora C, Rovira P, Richarte V, Garcia-Martinez I, Pagerols M, et al. Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. Mol Psychiatry. 2019.
23. Rabinowitz JA, Musci RJ, Milam AJ, Benke K, Uhl GR, Sisto DY, et al. The interplay between externalizing disorders polygenic risk scores and contextual factors on the development of marijuana use disorders. Drug and alcohol dependence. 2018;191:365-73.
24. Gurriaran X, Rodriguez-Lopez J, Florez G, Pereiro C, Fernandez JM, Farinas E, et al. Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. Genes Brain Behav. 2019;18(3):e12504.
25. Riglin L, Collishaw S, Thapar AK, Dalsgaard S, Langley K, Smith GD, et al. Association of Genetic Risk Variants With Attention-Deficit/Hyperactivity Disorder Trajectories in the General Population. JAMA Psychiatry. 2016;73(12):1285-92.

26. Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJ, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170(8):909-16.
27. Pedersen CB, Gøtzsche H, Møller JØ, Mortensen PB. The Danish Civil Registration System - A cohort of eight million persons. Dan Med Bull. 2006;53(4):441-9.
28. Norgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening Biobank. J Inherit Metab Dis. 2007;30(4):530-6.
29. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Health. 2011;39(7 Suppl):54-7.
30. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011;39(7 Suppl):30-3.
31. World Health Organization. International classification of diseases: Manual of the international statistical classification of diseases, injuries and causes of death (ICD-8). Geneva: WHO; 1967.
32. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva, Switzerland: World Health Organization; 1993.
33. Petersson F, Baadsgaard M, Thygesen LC. Danish registers on personal labour market affiliation. Scand J Public Health. 2011;39(7 Suppl):95-8.
34. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91.
35. Nagelkerke N. A note on a general definition of the coefficient of determination. Biometrika. 1991;78(3):691-2.
36. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012;36(3):214-24.
37. Molina BS, Hinshaw SP, Eugene Arnold L, Swanson JM, Pelham WE, Hechtman L, et al. Adolescent substance use in the multimodal treatment study of attention-deficit/hyperactivity disorder (ADHD) (MTA) as a function of childhood ADHD, random assignment to childhood treatments, and subsequent medication. Journal of the American Academy of Child and Adolescent Psychiatry. 2013;52(3):250-63.
38. Keller MC. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol Psychiatry. 2014;75(1):18-24.
39. Stata. Release 14. Statistical Software. College Station, TX: StataCorp, ; 2016.
40. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. Plos Genet. 2006;2(12):e190.
41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.
42. Shaun Purcell CC. Plink v1.90b3v 64-bit. v1.90b3v 64-bit ed2015.
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007;81(0002-9297; 0002-9297; 3):559-75.
44. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635-49.
45. Quinn P, Wigal S. Perceptions of girls and ADHD: results from a national survey. MedGenMed. 2004;6(2):2.

46. Madsen KB, Ravn MH, Arnfred J, Olsen J, Rask CU, Obel C. Characteristics of undiagnosed children with parent-reported ADHD behaviour. Eur Child Adolesc Psychiatry. 2018;27(2):149-58.
47. Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry. 2007;64(5):566-76.
48. Sorensen HJ, Manzardo AM, Knop J, Penick EC, Madarasz W, Nickel EJ, et al. The contribution of parental alcohol use disorders and other psychiatric illness to the risk of alcohol use disorders in the offspring. Alcoholism, clinical and experimental research. 2011;35(7):1315-20.
49. Toftdahl NG, Nordentoft M, Hjorthoj C. Prevalence of substance use disorders in psychiatric patients: a nationwide Danish population-based study. Soc Psychiatry Psychiatr Epidemiol. 2015.
50. Mohr-Jensen C, Vinkel Koch S, Briciet Lauritsen M, Steinhausen HC. The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. Eur Psychiatry. 2016;35:16-24.
51. Hansen SS, Munk-Jorgensen P, Guldbaek B, Solgard T, Lauszus KS, Albrechtsen N, et al. Psychoactive substance use diagnoses among psychiatric in-patients. Acta Psychiatr Scand. 2000;102(6):432-8.
52. Dalsgaard S, Humlum MK, Nielsen HS, Simonsen M. Common Danish standards in prescribing medication for children and adolescents with ADHD. Eur Child Adolesc Psychiatry. 2014;23(9):841-4.
53. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2018;47(1):226-35.
54. Wainschtein P, Deepti PJ, Yengo L, Zheng Z, Group. TAW, Consortium. T-OfPM, et al. Recovery of trait heritability from whole genome sequence data. bioRxiv. 2019.

TABLES AND FIGURES

Figure legends

Figure 1: Association between polygenic risk score for ADHD (PRS-ADHD) and any substance use disorders (SUD). Odds ratios and 95% confidence intervals for SUD are presented by deciles of the PRS-ADHD, in comparison to the lowest decile (reference). Analyses were adjusted for observation time, sex, age and calendar year at first ADHD diagnosis, and population stratification.

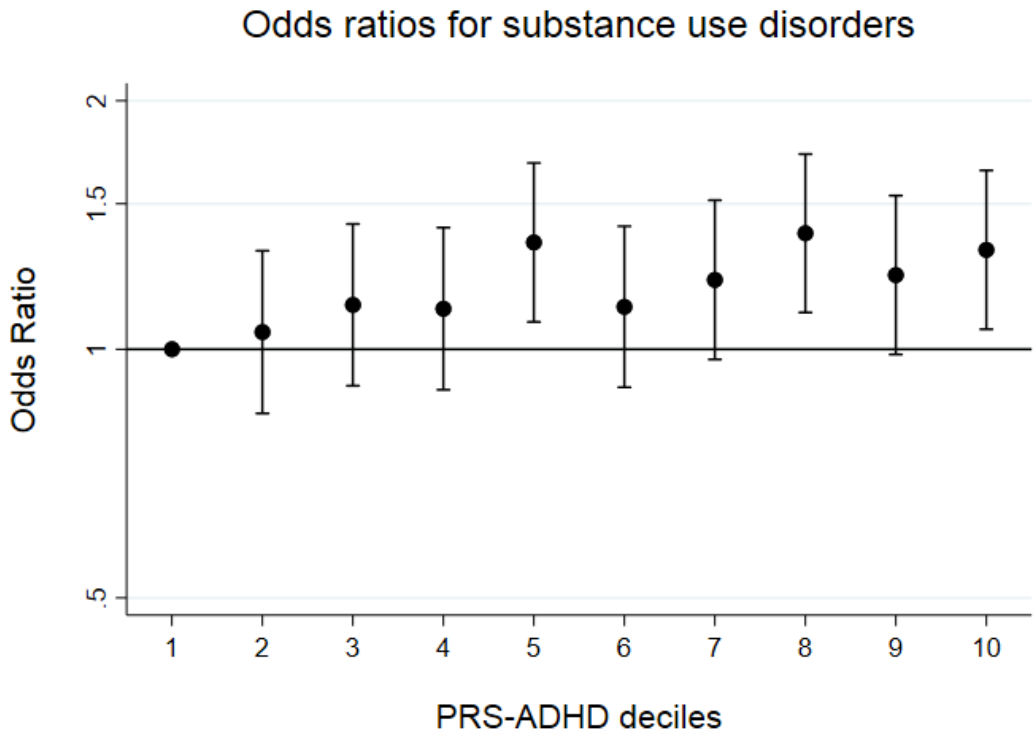
Figure 2: Forest plot of polygenic risk score for ADHD (PRS-ADHD) and other risk factors and their associations with any substance use disorder (SUD), SUD type, and SUD severity. The factors include: PRS-ADHD; coded in deciles but included continouosly and ranging from 0 to 1), sex (male vs. female), age at first ADHD-diagnosis (≥13 years vs. < 13 years), presence of comorbid oppositional defiant disorder/conduct disorder (ODD/CD), parental SUD, parental mental disorders, paternal income in the lowest quintile, and low maternal education (having compulsory education as the highest completed level of education). Associations were estimated as odds ratios for SUD, with 95% confidence intervals. Analyses were adjusted for observation time, sex, age and calendar year at first ADHD diagnosis (and in analysis of PRS-ADHD, also for the first four principal components to adjust for remaining population stratification).

Table 1: Patient- and family-related characteristics in individuals with and without substance use disorder in the study population of individuals with ADHD (N=13 116).

	Any SUD (n=2 368)	**No SUD** (n=10 748)	**All** (13 116)
Patient-related factors			
PRS-ADHD [€], mean (SD)	0.06 (0.99)	-0.01 (1.00)	0.00 (1.00)
Male sex, N (%)	1 731 (73.1)	7 949 (74.0)	9 680 (73.8)
ODD/CD, N (%)	558 (23.6)	1 263 (11.8)	1 821 (13.9)
Age at first ADHD ≥13 years [&], N (%)	1 865 (78.8)	3 695 (34.4)	5 560 (42.4)
Age at first ADHD, median (IQR)	18.6 (14.1-22.7)	10.5 (8.0-15.2)	11.5 (8.4-17.3)
Calendar year at first ADHD, median (IQR)	2009 (2006-2011)	2008 (2005-2010)	2008 (2005-2010)
Age at first ODD/CD, median (IQR)	16.2 (13-20) [$]	10.3 (7.6-14.6)	12.2 (8.2-16.4)
Age at first SUD, median (IQR)	19.4 (17.2-22.3)	-	-
Observation time (years), median (IQR)	13.6 (10.6-17.1)	7.7 (4.0-11.9)	9.0 (4.6-13.4)
Family-related factors			
Parental SUD, N (%)	460 (19.4)	1 468 (13.7)	1 928 (14.7)
Parental mental disorder, N (%)	868 (36.7)	3 719 (34.6)	4 587 (35.0)
Paternal income (low), N (%)	479 (20.2)	1 530 (14.2)	2 009 (15.3)
Maternal education (low), N (%)	1 071 (45.2)	3 325 (30.9)	4 396 (33.5)

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] The polygenic risk score is here included as a continuous normally distributed variable standardized within target-groups (defined by the 'leave-one-out'-approach for calculating the PRS-ADHD) in the final study population (n=13 116).
[$] Numbers were based on few individuals and rounded due to data regulations at Statistics Denmark.
[&] Age was dichotomized in order to make the effect size more comparable to the other risk factors.


Odds ratios for substance use disorders
Odds Ratio
.5
1
1.5
2
1
2
3
4
5
6
7
8
9
10
PRS-ADHD deciles

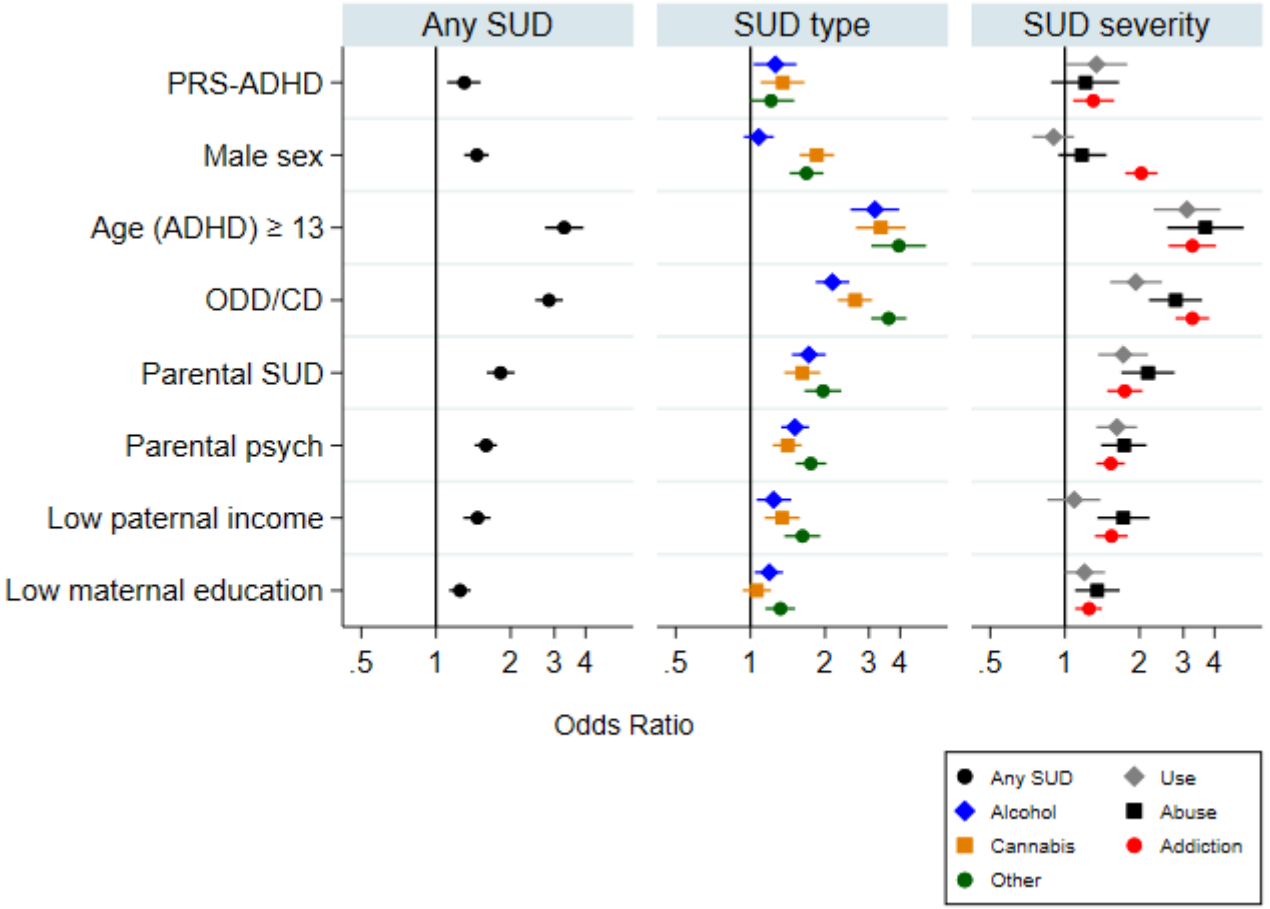

Any SUD
SUD type
SUD severity
PRS-ADHD
Male sex
Age (ADHD) ≥ 13
ODD/CD
Parental SUD
Parental psych
Low paternal income
Low maternal education
.5
1
2
3
4
Odds Ratio
Any SUD
Alcohol
Cannabis
Other
Use
Abuse
Addiction

Table 2: Associations between a range of risk factors and any SUD in individuals with ADHD (N=13 116), expressed as ORs, with 95% CIs. The proportion of variance in SUD explained by each risk factor is estimated by Nagelkerke's R^2 and a transformation to the liability scale.

Risk factors	OR (95% CI) [€]	$R^2_{Nagelkerke}$ (95% CI) [$] (%)	$R^2_{liability}$ (95% CI) [$] (%)
PRS-ADHD [§]	1.30 (1.11-1.51)	0.14 (0.02-0.38)	0.23 (0.03-0.60)
Male sex	1.46 (1.30-1.63)	0.61 (0.29-1.11)	0.96 (0.46-1.64)
Age at first ADHD ≥13 years [&]	3.28 (2.75-3.92)	2.26 (1.63-2.98)	3.57 (2.58-4.70)
ODD/CD	2.87 (2.53-3.26)	3.58 (2.71-4.51)	5.63 (4.29-7.08)
Parental SUD	1.83 (1.60-2.08)	1.06 (0.63-1.61)	1.69 (1.01-0.03)
Parental mental disorder	1.59 (1.43-1.76)	0.99 (0.58-1.51)	1.58 (0.92-2.40)
Paternal income (low)	1.46 (1.29-1.66)	0.47 (0.20-0.85)	0.75 (0.31-1.35)
Maternal education (low)	1.25 (1.13-1.38)	0.26 (0.07-0.56)	0.41 (0.12-0.89)

Abbreviations: SUD: substance use disorder. ADHD: attention-deficit/hyperactivity disorder. OR: odds ratio. CI: confidence interval. PRS: polygenic risk score. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] Analyses were adjusted for observation time, sex and age and calendar year at first ADHD diagnosis. For the PRS-ADHD, the model additionally included the first four principal components to adjust for remaining population stratification.
[$] All estimates were based on log likelihood estimates from logistic regression analyses comparing models with and without the factor of interest. All models were adjusted for observation time, calendar year at first ADHD diagnosis, the first four principal components. Analyses were not adjusted for sex and age at first ADHD diagnosis for a meaningful model comparability. Nagelkerke's R^2 was estimated and transformed to the liability scale as proposed by Lee et al. (36) assuming a population prevalence of SUD of 40%.(22, 37) R^2 measures were accompanied by non-parametric 95% CIs using bootstrap resampling with 10 000 replications.
[§] The ADHD-PRS was categorized into deciles but included continuously yielding ORs corresponding to comparison of the highest decile to the lowest.
[&] Age were dichotomized in order to make the effect size more comparable to the other risk factors.

Table 3: Associations between the PRS-ADHD and different types of SUD and of different severity levels of SUD in individuals with ADHD (N=13 116), expressed as ORs, with 95% CIs, in different adjustment models.

		Types of SUD[§]			Severity levels of SUD[&]		
The different covariates included in the adjustment models	Any SUD (n=2 368)	Alcohol (n=1 207)	Cannabis (n=1 106)	Other illicit drugs (n=1 071)	Use (n=522)	Abuse (n=414)	Addiction (n=1 432)
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
M1: PRS-ADHD[€], minimal adjustment[$]	1.30 (1.11-1.51)	1.26 (1.04-1.53)	1.34 (1.10-1.64)	1.21 (0.99-1.50)	1.36 (1.02-1.80)	1.21 (0.88-1.65)	1.30 (1.07-1.57)
M2: M1 + ODD/CD	1.25 (1.07-1.47)	1.23 (1.01-1.50)	1.30 (1.06-1.60)	1.17 (0.95-1.45)	1.34 (1.01-1.77)	1.19 (0.87-1.63)	1.27 (1.05-1.54)
M3: M2 + Parental SUD	1.23 (1.06-1.44)	1.21 (0.99-1.47)	1.29 (1.05-1.58)	1.15 (0.93-1.42)	1.32 (0.99-1.75)	1.18 (0.86-1.62)	1.25 (1.02-1.51)
M4: M3 + Parental psych	1.23 (1.05-1.44)	1.20 (0.99-1.47)	1.28 (1.05-1.57)	1.14 (0.92-1.41)	1.31 (0.99-1.73)	1.17 (0.85-1.61)	1.24 (1.02-1.51)
M5: M4 + Paternal income	1.22 (1.04-1.42)	1.20 (0.99-1.47)	1.27 (1.04-1.56)	1.12 (0.91-1.39)	1.32 (0.99-1.75)	1.14 (0.83-1.57)	1.22 (1.00-1.48)
M6: M5 + Maternal education	1.21 (1.03-1.41)	1.20 (0.98-1.46)	1.28 (1.04-1.57)	1.11 (0.90-1.38)	1.31 (0.99-1.74)	1.13 (0.82-1.56)	1.21 (1.00-1.48)

Abbreviations: SUD: substance use disorder. ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. OR: odds ratio. CI: confidence interval. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] The PRS-ADHD was categorized into deciles but included continuously yielding ORs corresponding to comparison of the highest decile to the lowest.
[$] Adjusted for observation time, sex and age and calendar year at first ADHD diagnosis, and the first four principal components to adjust for remaining population stratification.
[§] Not mutually exclusive groups. Individuals with ADHD and several types of SUD are included in more than one group.
[&] Hierarchically ordered into mutually exclusive groups, with the following total sample sizes: Use (N=11 270), abuse (N=11 162), and addiction (N=12 180). In all three analyses, the reference category includes individuals with no registered diagnosis of use, abuse or addiction.

SUPPLEMENTARY MATERIAL

Calculation of polygenic risk scores for ADHD (PRS-ADHD)

PRS-ADHD were derived from summary statistics in the meta-analysis in recently published genome-wide-association study (GWAS),(1) applied in the iPSYCH2012 case-cohort sample.(2) Polygenic risk scores were based on summary statistics from the GWAS meta-analysis including both the iPSYCH2012 case-cohort sample (14 584 cases and 22 492 controls) as well as the 10 samples with European ancestry from the Psychiatric Genomics Consortium (PGC) (4 578 cases and 11 912 controls). The specific samples and genotyping procedure are described in detail elsewhere.(1) To include information from the iPSYCH2012 case-cohort sample in the discovery sample, PRS were calculated in a so-called 'leave-one-out' approach: Individuals in the iPSYCH2012 case-cohort sample were split into five target groups. This was done so that the discovery and target datasets were non-overlapping by assigning the 23 waves into five groups, distributing the total number of individuals, the number of ADHD cases, and birth years as equal as possible across the five target groups. Polygenic risk scores were then estimated using the ricopili pipeline(3) for each of the five target samples. Risk alleles were identified in a GWAS on potential risk alleles for ADHD in a discovery sample consisting of the 10 PGC European samples and iPSYCH sample leaving out the target group. Scores were calculated as the sum of the ADHD risk alleles identified from the GWAS meta-analysis as SNPs passing a specific significance threshold and weighted by the effect sizes identified in the discovery sample. To assign a PRS to all individuals in the iPSYCH sample, this procedure was repeated for each of the five target groups. PRS were calculated for each of the following 10 significance thresholds: 0.00000005, 0.000001, 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, 1. Within target datasets, only SNPs intersecting across all genotyping waves with an info score of at least 0.6 and a MAF of at least 0.01 were used when calculating PRS. Following Demontis et al.,(1) for the main analyses we included SNPs at a p-value threshold at 0.2, optimizing the variance explained according to ADHD case status. For this chosen threshold the number of SNPs included in the PRS calculation ranged from 61 179 (target group 1) to

61 637 (target group 4). Additionally, 10 PRSs, one for each p-value threshold, were similarly calculated excluding the PGC European samples in the discovery sample. Finally, we standardized the PRS-ADHD within the five target samples and generated deciles.

Method for restriction to European ancestry and unrelated individuals

The training of the PRS-ADHD was based entirely on European samples and thus may be biased in populations of non-European ancestry.(4) We initially restricted our target dataset to ADHD cases with both parents born in Denmark to include parental information from the registers. Still, there might be remaining ancestry outliers, and we therefore further restricted our study population to individuals with European ancestry. This was done following a standard principal component analysis (PCA)-based approach similar to what has been done in the ADHD GWAS meta-analysis.(1)In order to identify population substructure and ancestry, we used principal components (PCs) from a PCA conducted for 81 620 individuals using EIGENSOFT version 6.1.4(5) on a relatedness-pruned set of individuals and a subset of SNPs. Further description for this PCA procedure as well as a visual inspection of the first principal components (PCs) showing clear patterns of population structure can be found in Pedersen et al. 2018(6).

We calculated a mean and standard deviation of each of the first three principal components in a population of individuals in the entire iPSYCH sample with all four grandparents born in Denmark. From the estimated parameters, we removed all with values outside a three-dimensional ellipsoid centered at the estimated means and with an ellipsoid at a distance of six standard deviations from the mean in each direction.

Next, we removed genetically close related individuals by identifying individuals with an estimated relatedness coefficient of $\hat{\pi} > 0.2$ in

plink version 1.9,(7-9) and kept one of the related individuals, including as many unrelated individuals as possible in our final study population.

Last, we calculated new PCs based on our study population after restriction to European ancestry and without close relationships to check for remaining population structure. The PCA was based on an LD-pruned set of SNPs (from 496,370 to 25,085, conducted in plink version 1.9 with the following parameters window-size = 1500 kb, step-size = 500, $r^2 = 0.02$). After further excluding 9 SNPs with a high loading on the first PC, causing approximately 50 outliers from genotyping wave number 23 and therefore likely to be genotyping errors, a PCA resulted in PCs indicating no remaining population stratification. Furthermore, we tested correlations between the calculated PCs and the PRS-ADHD as well as our outcome of SUD, and found only week evidence of an association with the first 10 PCs. We decided to include the first four PCs in the main analyses, and to evaluate the sensitivity of results, we repeated the main analysis including no PCs and 10 PCs, respectively, in the analysis.

Additional sensitivity analyses

We conducted several additional sensitivity analyses, using different model specifications and study populations, to check robustness of our main estimate for the association between PRS-ADHD and SUD. These analyses are described in the following and the results are presented in Table S8.

First, we calculated PRS-ADHD excluding the PGC samples in the discovery sample, i.e. using summary statistics entirely based on the iPSYCH sample. Next, we investigated the impact of the number of principal components included in the model by repeating the analysis including no and the first 10 PCs, respectively. In the main analysis, we initially restricted our cohort to individuals born in Denmark with

Danish-born parents (initial ADHD cohort), and then further restricted to genetically unrelated individuals and individuals with European ancestry. This was done in accordance with pertinent research(10, 11) to ensure that the PRS-ADHD were applied in a homogeneous Danish population with non-European ancestry. However, to explore robustness and generalizability of results to the less restrictive initial ADHD cohort, we explored the association, without excluding genetically related individuals and individuals with a 3rd generation non-European ancestry. Third, to evaluate the impact of the timing of exposure and outcome, analyses were also repeated for ADHD sub-cohorts ensuring that the ADHD diagnosis was not given after the first SUD diagnosis. This was done by restricting the cohort to individuals with no SUD diagnosis before first ADHD diagnosis and by restricting the cohort to individuals with the first ADHD diagnosis before age 13. Finally, we restricted to ADHD cases with 10 years observation time after age 13 years. This restriction included the older birth cohorts, but ensured same time at risk of being diagnosed with SUD.

As demonstrated in Table S8, these changes to the model and cohort restrictions barely changed the estimates, as compared to our main analysis.

Table S1: Study population, applying different cohort restrictions

Inclusion criterion	**Excluded, n (%)**	**Included, n**
All sampled as ADHD-cases in iPSYCH2012[€]		18 726
Individuals with biological material available in the Danish Neonatal Screening Biobank	891 (4.8)	17 835
Passing genetic quality control	586 (3.3)	17 249
Non-missing information on PRS-ADHD	195 (1.1)	17 054
Parents born in Denmark	2202 (12.9)	14 852
Ignoring ADHD diagnoses before age 3	84 (0.6)	14 768
Born before Dec 31, 2003 (to enable follow-up from age 13)	662 (4.5)	14 106
Alive, not emigrated from Denmark before age 13	7 (0.0)	14 099
No SUD before age 13	31 (0.0)	14 068
Restricting to unrelated individuals	756 (5.4)	13 312
Restricting to European ancestry	196 (1.5)	**13 116**

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder.
[€] Inclusion criteria used in the iPSYCH-sample: Individuals born in Denmark between May 1, 1981 – Dec 31, 2005, known mother, ADHD defined as ICD-10 diagnosis F90.0 after age 1 and until Dec 31, 2012.

Table S2: ICD-8 and ICD-10 diagnostic codes used for different disease categories[$]

Disease	ICD-8	ICD-10
ADHD	308.01[§]	F90.0, F98.8[§]
Any Substance use disorder (SUD)		F10-F19
Types of SUD		
Alcohol		F10
Cannabis		F12
Other illicit drugs		F11, 13-16, 18-19
Severity levels of SUD		
Use		F1x.0
Abuse		F1x.1, F1x.8, F1x.9
Addiction		F1x.2, F1x.3, F1x.4, F1x.5, F1x.6, F1x.7
ODD/CD	308.03-06	F90.1, F91
Parental mental disorders	290-315	F00-F99
Parental SUD	291.x9, 294.39, 303.x9, 303.20, 303.28, 303.90, 304.x9	F10-F19

Abbreviations: ICD-8 and ICD-10: The International Classification of Diseases, 8th and 10th edition. ADHD: attention-deficit/hyperactivity disorder. SUD: substance use disorder. ODD/CD: oppositional defiant disorder/conduct disorder.

[$] Identical with the definitions previously published by Pedersen CB, Mors O, Bertelsen A, Waltoft BL, Agerbo E, McGrath JJ et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA Psychiatry 2014; 71(5): 573-581.

[§] The codes 308.01 and F98.8 were not part of the definition of ADHD in the iPSYCH2012 cohort, which only included F90.0.

Table S3: Associations between a range of risk factors and different types of SUD and of different severity levels of SUD in individuals with ADHD (N=13 116), expressed as ORs with 95% CIs. [€] *Numbers equal the estimates presented in Figure 2.*

		Types of SUD[§]			Severity levels of SUD[&]		
	Any SUD (n=2 363)	Alcohol (n=1 204)	Cannabis (n=1 103)	Other illicit drugs (n=1 071)	Use (n=520)	Abuse (n=414)	Addiction (n=1 429)
Risk factors	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
PRS-ADHD[$]	1.30 (1.11-1.51)	1.26 (1.04-1.53)	1.34 (1.10-1.64)	1.21 (0.98-1.50)	1.36 (1.02-1.80)	1.21 (0.88-1.65)	1.30 (1.07-1.57)
Male vs. female sex	1.46 (1.30-1.63)	1.08 (0.94-1.24)	1.84 (1.57-2.16)	1.67 (1.43-1.96)	0.90 (0.74-1.09)	1.17 (0.93-1.46)	2.02 (1.74-2.34)
Age at first ADHD ≥13 years	3.28 (2.75-3.92)	3.17 (2.53-3.97)	3.33 (2.64-4.19)	3.95 (3.07-5.09)	3.12 (2.29-4.25)	3.68 (2.58-5.23)	3.24 (2.59-4.04)
ODD/CD	2.87 (2.53-3.26)	2.16 (1.84-2.52)	2.65 (2.26-3.10)	3.63 (3.08-4.27)	1.94 (1.53-2.47)	2.79 (2.18-3.57)	3.29 (2.81-3.84)
Parental SUD	1.83 (1.60-2.08)	1.73 (1.47-2.02)	1.62 (1.38-1.92)	1.96 (1.66-2.32)	1.72 (1.37-2.17)	2.16 (1.69-2.76)	1.76 (1.49-2.06)
Parental mental disorder	1.59 (1.43-1.76)	1.51 (1.33-1.72)	1.41 (1.23-1.62)	1.74 (1.52-2.01)	1.62 (1.35-1.96)	1.72 (1.40-2.13)	1.53 (1.34-1.74)
Paternal income (low)	1.46 (1.29-1.66)	1.24 (1.06-1.45)	1.34 (1.14-1.58)	1.61 (1.37-1.90)	1.08 (0.85-1.39)	1.72 (1.35-2.19)	1.54 (1.32-1.79)
Maternal education (low)	1.25 (1.13-1.38)	1.19 (1.05-1.35)	1.06 (0.93-1.21)	1.32 (1.15-1.51)	1.21 (1.00-1.46)	1.35 (1.10-1.66)	1.25 (1.10-1.41)

Abbreviations: SUD: substance use disorder. ADHD: attention-deficit/hyperactivity disorder. OR: odds ratio. CI: confidence interval. PRS: polygenic risk score. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] All analyses were adjusted for observation time, sex and age and calendar year at first ADHD diagnosis, and the first four principal components to adjust for remaining population stratification.

[$] The PRS-ADHD was included continuously in the models, and estimates correspond to a comparison of the highest vs. lowest decile of PRS-ADHD.

[§] Not mutually exclusive groups. Individuals with ADHD and several types of SUD are included in more than one group.

[&] Hierarchically ordered into mutually exclusive groups, with the following total sample sizes: Use (N=11 270), abuse (N=11 162), and addiction (N=12 180). In all three analyses, the reference category includes individuals with no registered diagnosis of use, abuse or addiction.

Table S4: Mean PRS-ADHD differences in risk factors for SUD

Risk factors	Mean difference in PRS-ADHD [€] (95% CI)	p [$]
Male vs. female sex	-0.01 (-0.05-0.03)	0.49
Age at first ADHD ≥13 years	-0.03 (-0.06-0.01)	0.15
ODD/CD	0.14 (0.09-0.18)	<0.0001
Parental SUD	0.08 (0.04-0.13)	<0.001
Parental mental disorder	0.07 (0.04-0.11)	<0.0001
Paternal income (low)	0.17 (0.12-0.22)	<0.0001
Maternal education (low)	0.14 (0.10-0.18)	<0.0001

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. CI: confidence interval. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] The PRS-ADHD was here included as a continuous normally distributed variable standardized within target-groups (defined by the 'leave-one-out' approach for calculating the PRS-ADHD) in the final study population (n=13 116).

[$] The p-values should be evaluated according to a Bonferroni-corrected p-value threshold of 0.05/7=0.007.

Table S5: Interactions between the PRS-ADHD and other risk factors for SUD

Risk factors	**p-value [$] for the interaction with PRS-ADHD [€]**
Male vs. female sex	0.07
Age at first ADHD ≥13 years	0.39
ODD/CD	0.07
Parental SUD	0.31
Parental mental disorder	0.06
Paternal income (low)	0.56
Maternal education (low)	0.33

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. CI: confidence interval. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] The PRS-ADHD was here included as a continuous normally distributed variable in the final study population (n=13 116).

The models are adjusted for observation time, and further include the following covariates: sex and age and calendar year at first ADHD diagnosis, and the first four principal components to adjust for remaining population stratification. All interactions between covariates and the two interaction variables of interest were included in the models.

[$] The p-values should be evaluated according to a Bonferroni-corrected p-value threshold of 0.05/7=0.007.

Table S6: Sex-specific associations between the PRS-ADHD and SUD.[$] expressed as ORs with 95% CI,.[€] for types and severities of SUD, in males (n=9 680) and females (n=3 436).

		Types of SUD[§]			Severity levels of SUD[&]		
	Any SUD	Alcohol	Cannabis	Other illicit drugs	Use	Abuse	Addiction
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
All	1.30 (1.11-1.51)	1.26 (1.04-1.53)	1.34 (1.10-1.64)	1.21 (0.98-1.50)	1.36 (1.02-1.80)	1.21 (0.88-1.65)	1.30 (1.07-1.57)
Males	1.18 (0.98-1.42)	1.26 (0.99-1.59)	1.26 (1.01-1.60)	1.06 (0.83-1.35)	1.15 (0.82-1.63)	1.28 (0.87-1.87)	1.16 (0.93-1.45)
Females	1.59 (1.19-2.12)	1.22 (0.86-1.73)	1.58 (1.03-2.40)	1.78 (1.18-2.69)	1.77 (1.09-2.85)	1.05 (0.60-1.84)	1.76 (1.20-2.58)

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. OR: odds ratio. CI: confidence interval.
[€] All analyses were adjusted for observation time, and age and calendar year at first ADHD diagnosis, and the first four principal components to adjust for remaining population stratification.
[$] The polygenic risk score for ADHD (PRS-ADHD) was included continuously in the models, and estimates correspond to a comparison of the highest vs. lowest decile of PRS-ADHD.
[§] Not mutually exclusive groups. Individuals with ADHD and several types of SUD are included in more than one group.
[&] Hierarchically ordered into mutually exclusive groups. In all three analyses, the reference category includes individuals with no registered diagnosis of use, abuse or addiction.

Table S7: Associations between PRS-ADHD and SUD for 10 p-value thresholds for SNPs included to calculate the PRS-ADHD. ORs[$] and 95% CIs are presented for individuals in the study population (n= 13 116). The PRS-ADHD used in the main analyses was based on the significance threshold of 0.2 and is highlighted in grey.

PRS score	Significance threshold (max. p-value)	Number of SNPs [€]	OR (95%CI)
1	0.00000005	7	1.02 (0.87-1.19)
2	0.000001	26	1.05 (0.90-1.22)
3	0.0001	319	1.07 (0.92-1.25)
4	0.001	1477	1.16 (0.99-1.35)
5	0.01	7676	1.25 (1.07-1.46)
6	0.05	24 004	1.38 (1.19-1.61)
7	0.1	38 726	1.36 (1.17-1.59)
8	0.2	61 401	1.30 (1.11-1.51)
9	0.5	106 739	1.34 (1.15-1.56)
10	1	141 684	1.33 (1.14-1.55)

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. SNP: single nucleotide polymorphism. OR: odds ratio. CI: confidence interval.

[€] The number presented is an average of the number of SNPs identified in the training data sets corresponding to the five target groups. For PRS-ADHD score number 8, the number of SNPs varies from 61 179 (group 1) to 61 637 (group 3).

[$] Analyses are adjusted for observation time, sex, and age and calendar year at first ADHD diagnosis and the first four principal components to adjust for remaining population stratification.

Table S8: Proportion of variance explained by a range of risk factors SUD in individuals with ADHD (N=13 116), estimated by Nagelkerke's R^2 and a transformation to the liability scale for different assumed population prevalences of SUD.

Risk factors	$R^2_{Nagelkerke}$ (95% CI) [€] (%)	$R^2_{liability}$ (95% CI) [€] (%)		
Population prevalence of SUD (k)		k=35%	k=40%	k=45%
PRS-ADHD [$]	0.14 (0.02-0.38)	0.22 (0.03-0.58)	0.23 (0.03-0.60)	0.23 (0.03-0.60)
Male sex	0.61 (0.29-1.11)	0.94 (0.45-1.64)	0.96 (0.46-1.64)	0.98 (0.47-1.69)
Age at first ADHD ≥13 years	2.26 (1.63-2.98)	3.49 (2.52-4.61)	3.57 (2.58-4.70)	3.61 (2.61-4.81)
ODD/CD	3.58 (2.71-4.51)	5.51 (4.23-6.95)	5.63 (4.29-7.08)	5.70 (4.39-7.21)
Parental SUD	1.06 (0.63-1.61)	1.65 (0.98-2.49)	1.69 (1.01-0.03)	1.71 (1.00-2.57)
Parental mental disorder	0.99 (0.58-1.51)	1.54 (0.90-2.35)	1.58 (0.92-2.40)	1.60 (0.93-2.42)
Paternal income (low)	0.47 (0.20-0.85)	0.74 (0.31-1.34)	0.75 (0.31-1.35)	0.76 (0.32-1.38)
Maternal education (low)	0.26 (0.07-0.56)	0.40 (0.11-0.89)	0.41 (0.12-0.89)	0.42 (0.12-0.90)

Abbreviations: SUD: substance use disorder. ADHD: attention-deficit/hyperactivity disorder. OR: odds ratio. CI: confidence interval. PRS: polygenic risk score. ODD/CD: oppositional defiant disorder/conduct disorder.

[€] All estimates were based on log likelihood estimates from logistic regression analyses comparing models with and without the factor of interest. All models were adjusted for observation time, calendar year at first ADHD diagnosis, the first four principal components. Analyses were not adjusted for sex and age at first ADHD diagnosis for a meaningful model comparability. Nagelkerke's R^2 was estimated and transformed to the liability scale as proposed by Lee et al. (12) assuming a population prevalence of SUD of 35%, 40%, and 45%, respectively.(13, 14) R^2 measures were accompanied by non-parametric 95% CIs using bootstrap resampling with 10 000 replications.

[$] The ADHD-PRS was categorized into deciles but included continuously yielding ORs corresponding to comparison of the highest decile to the lowest.

Table S9: Additional sensitivity analyses. All estimates are based on a logistic regression with the PRS-ADHD as the exposure and any SUD as the outcome and adjustment similar to the main analysis (n=13 116), if not otherwise specified.

Setting	Total, n	SUD, n (%)	OR (95% CI)
Main analysis[€]	13 116	2 368 (18.1%)	1.30 (1.11-1.51)
PRS-ADHD excluding PGC in the discovery sample	13 116	2 368 (18.1%)	1.30 (1.12-1.52)
Including no principal components	13 116	2 368 (18.1%)	1.30 (1.11-1.51)
Including the first 10 principal components	13 116	2 368 (18.1%)	1.30 (1.11-1.51)
No restriction to genetically unrelated individuals or European ancestry[$]	14 068	2 530 (18.0%)	1.33 (1.14-1.54)
No PCA-based restriction to European ancestry[$]	13 312	2 400 (18.0%)	1.30 (1.11-1.51)
Restricting to ADHD cases with no SUD diagnosis before first ADHD diagnosis	12 309	1 561 (12.7%)	1.28 (1.08-1.53)
Restricting to ADHD cases diagnosed with ADHD before age 13	7 556	503 (6.7%)	1.43 (1.06-1.93)
Restricting to ADHD cases with a fixed observation time of 10 years after age 13, i.e. restricting to the oldest ADHD cases with at least 10 years of follow up and ignoring SUD diagnoses after age 23	5 687	1 427 (25.1%)	1.31 (1.08-1.58)

Abbreviations: ADHD: attention-deficit/hyperactivity disorder. PRS: polygenic risk score. SUD: substance use disorder. OR: odds ratio. CI: confidence interval. PGC: Psychiatric Genomics Consortium.

[€] Adjusted for observation time, sex, age and calendar year at first ADHD diagnosis, and the first four principal components to adjust for remaining population stratification. The estimates for the polygenic risk score (PRS-ADHD) correspond to a comparison of the highest vs lowest decile of PRS for ADHD.

[$] Adjusted for population stratification by including the first four principal components from the original principal component analysis.

References

1. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51(1):63-75.
2. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Baekvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2017.
3. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421-7.
4. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635-49.
5. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. Plos Genet. 2006;2(12):e190.
6. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Baekvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2018;23(1):6-14.
7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.
8. Shaun Purcell CC. Plink v1.90b3v 64-bit. v1.90b3v 64-bit ed2015.
9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007;81(0002-9297; 0002-9297; 3):559-75.
10. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91.
11. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. Nat Commun. 2019;10(1):333.
12. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012;36(3):214-24.
13. Soler Artigas M, Sanchez-Mora C, Rovira P, Richarte V, Garcia-Martinez I, Pagerols M, et al. Attention-deficit/hyperactivity disorder and lifetime cannabis use: genetic overlap and causality. Mol Psychiatry. 2019.
14. Molina BS, Hinshaw SP, Eugene Arnold L, Swanson JM, Pelham WE, Hechtman L, et al. Adolescent substance use in the multimodal treatment study of attention-deficit/hyperactivity disorder (ADHD) (MTA) as a function of childhood ADHD, random assignment to childhood treatments, and subsequent medication. Journal of the American Academy of Child and Adolescent Psychiatry. 2013;52(3):250-63.